

JUL 12 1990

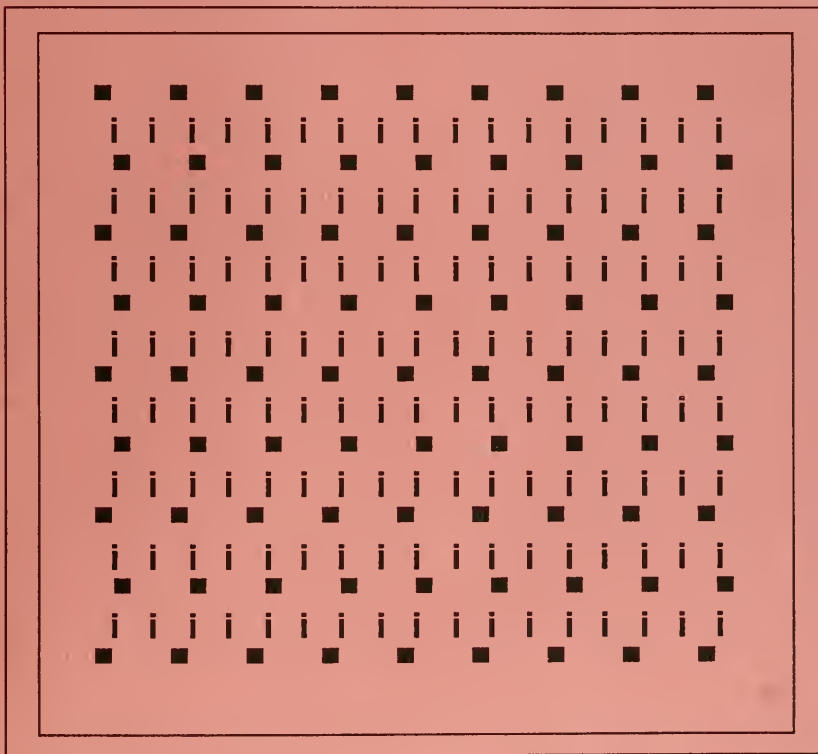
IASSIST

Q U A R T E R L Y

VOLUME 13

Fall/Winter 1989

NUMBER 3/4



IASSIST

QUARTERLY

The **IASSIST QUARTERLY** represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The **QUARTERLY** reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of **IASSIST**.

Information for Authors

The **QUARTERLY** is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press *Manual of Style* or Kate L. Turabian's *Manual for Writers*. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". *Journal of the American Society for Information Science* 30(2):77-82, March 1979. If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor: Walter Piovesan, Research Data Library, W.A.C. Bennett Library, Simon Fraser University, Burnaby, B.C., V5A 1S6 CANADA. (604) 291-4349 E-Mail: USERDLIB@SFU.BITNET. Book reviews should be submitted in duplicate to the Book Review Editor: Daniel Tsang, Main Library, University of California P.O. Box 19557, Irvine, California 92713 USA. (714) 856-4978 E-Mail: DTSANG@ORION.CF.UCL.EDU

Title: Newsletter - International Association for
Social Science Information Service and
Technology
ISSN - United States: 0739-1137 Copyright 1985 by
IASSIST. All rights reserved.

C O N T E N T S

Volume 13

Number 3/4

Fall/Winter 1989

FEATURES

- 4** **The record group concept in contemporary archives**
by P.A. Alsberg
- 7** **National archives and electronic records; where are we going**
by K.S. Gavrel
- 10** **Data in the natural and exact sciences and in the social sciences: a comparative study.**
by E. Hoffman
- 13** **Policy-oriented research in Britain: some institutional mechanisms**
by R. Barnes
- 16** **U.S. Bureau of the Census data dissemination programs**
by W.P. Butz
- 21** **Data from the Central Bureau of Statistics to the social science community: the Norwegian experience**
by B. Henriksen
- 24** **The role of the Hungarian Central Statistical Office in providing social science data**
By G. Vukovich

The Record Group Concept In Contemporary Archives

by Paul A. Alsberg

*Presented at the IFDO/IASSIST 89 Conference
held in Jerusalem, Israel, May 15-18, 1989*

All professional definitions of the term "Archives" use the concept of "an organic whole of documents growing out of the activities of a physical or legal body". This concept is reflected in the basic archival principle of "respect des fonds" or "principle of provenance". In translating this theoretical approach into the practical preservation of the "organic whole" of an administrative body two questions have to be answered:

1. How can we define a "body"? Are we speaking about the preservation of the most concise or the most comprehensive administrative organism?
2. By what characteristics can we decide if an administrative organism is considered to be complete by itself?

Johannes Papritz, at the International Congress on Archives in Brussels in 1964¹ related four criteria to the Soviet-Russian theoretical approach in defining the independent status of an administrative unit:

- 1) a legal basis of the unit, its tasks and competences; 2) an independent budget; 3) an establishment list of its own; 4) an official seal or stamp.

These four criteria may perhaps be considered to be a convenient basis for the comprehensive organism, but for the most concise body to be preserved as a whole it is more difficult to establish the criteria, and as archivists we try to find pragmatic solutions by establishing record groups and sub-groups.

However, the problem is much more complex because of the dynamics of modern administration and the frequent changes in organizations and administrative patterns. Some 20 years ago a number of papers were published advocating the abandonment of the Record Group Concept in Modern Records in favor of basing archival arrangement and inventorying on "Series"; this means the original record-keeping units as the "Whole" to be preserved and not the administrative bodies.

This thesis was put forward, i.e. by an Australian Archivist Peter J. Scott in a paper published in 1966². Scott based his approach on difficulties encountered with record-series created by one agency and later incorporated into another agency and transferred to the archives by this second or perhaps even a third agency. The second difficulty arose, according to Scott, by the practice of using the record groups for determining the

shelf-location. Let us answer immediately that this question of shelf-location was once considered to be of major importance, but already at the beginning of this century the Archivist of Danzig, Max Baer, propagated the solution of reconstructing the original order on paper only and storing and shelving the records according to administrative considerations³. The principle of provenance does not mean to arrange the records physically according to their original order.

Another call for the revision of the Record Group Concept in the National Archives of the U.S.A. was published the same year by Mario D. Fenyo⁴. He published a critical analysis, excellently written, of the concept and the history of the definition used in the National Archives. I have to confess that even 23 years ago, when first reading the paper, I was not convinced by the critique. On the contrary, I learned from the minutes of the deliberation of the committee appointed by the Archivist of the United States in 1940 and charged with formulating a policy for arranging and inventorying archives, published by Fenyo, how understanding and not dogmatic the first generation of archivists of the U.S. was. They learned the lessons from the European experience and adopted a system of their own. Theory is an end by itself, and Dr. Buck - the second Archivist of the United States - and his colleagues tried to combine theoretical and pragmatic solutions.

We have, therefore, to answer four main questions:

- a) Is the "record group" a concept in the field of archival management, or is the "record group" a concept in the sphere of archival theory?
- b) What purpose does the "record group" serve theoretically and administratively?
- c) What are the typical structures of modern governmental records?
- d) In what way do these structures influence the formation of "record groups"?

The basic answers to our first two questions were already given by the National Archives of the U.S.A. in 1941 and published in Staff Information Circular No. 15 in July 1950, when defining a record group as "a major archival unit established somewhat arbitrarily with due regard to the principle of provenance and to the desirability of

making the unit of convenient size and character for the work of arrangement and description and for the publication of inventories”.

This definition makes it clear that the “record group” as a concept belongs to archival management and not so much to archival theory. It should be stressed, however, that the “record group” is established - as said in the definition - in accordance with principles based on archival theory. In order to clarify what was apparently meant by the word in the definition “with due respect to the principle of provenance”, we have to dwell on these theoretical principles.

Three different frameworks form the foundations of the principle of provenance and can be the basis for a “record group”.

The first: The system for record-keeping, one single registry or one filing scheme. The second framework can be the essential function which an administration has to fulfill. This means the purpose for which an administrative body has been established. The third framework can be an administrative one. Generally it was considered that this administrative framework is the main basis for establishing a “record group” in accordance with the principle of provenance. This was certainly right at a time when one comprehensive administrative unit as a whole Ministry or, in the American administration, a whole Department or Bureau, had one single central registry. As a Ministry or Bureau had a well defined framework of functions, interdependent and interwoven, it had one single central registry. This assumption meant that there were in reality identical frameworks for record-keeping, administration and function.

In modern records, accumulated during the last 50 years - and this process is now accelerated by new technology and the ever increasing bureaucracy - we have quite different organizational set-ups.

Functions are more stable than the organizational framework. Political reasons bring about the establishment of new administrative units and even split up functional units and define in a more exact and sometimes sophisticated way two or three functions which were previously considered to be one organic whole. The disappearance or cessation of a function have no bearing on discussion, as the relevant series or “record group” simply cease to exist and are considered “closed”. But in addition to the splitting up of a function there is in the modern administration a much bigger differentiation in the keeping of records and nearly no modern office has any longer a central registry or central filing rooms. Today it would be anachronistic to think that the 3 different frameworks of record keeping, function and administration could jointly be the basis for the establishment of modern “record groups”. We have to admit, therefore, that each of these frameworks separately may form the basis of a “record group”.

In order to avoid misunderstandings I would like to differentiate between four types of “record groups”: 2 types

of organic and 2 types of artificial “record groups”. An organic record group could be the former type of one unit with one registry or one filing system pertaining mainly to one essential function. This function will generally be defined in one central law or one basic administrative decision, with all addition, corrections, regulations and ordinances. This means one indivisible record cell. The other type could be the organic record group of an administrative unit which has more than one record-keeping cell, different registries pertaining to different aspects of one or more laws.

I would like to stress that this concept based on record cells pertains to the essential functions only and excludes the auxiliary and housekeeping functions which have no bearing on the establishment of record groups. You may have one registry system for the main essential function and two or more additional systems for the auxiliary functions and we would still speak about a record group of one cell, as the auxiliary functions are not relevant to our analysis. Housekeeping records, even if they are not disposed of, are in any case of minor importance and no record group should be organized or established for them, besides the preservation of the necessary files of documentation on the organization as such.

In every record group one may find different chronological layers, strata based on simple cut-off periods or changes in the technical filing methods. As long as there are no basic changes of functions or authority these are different layers within one system.

Within this second type of organic group of more than one cell one may even have different chronological layers as every cell may have its own system, its own cut-offs and its own development.

There are also two types of artificial record groups. The first is artificial only in the way its organic cells are brought together under one roof for the purpose of archival management. Small archival units which are of one of the first two types, but too small to appear as a major unit in an archival institution, may be grouped together if they have common characteristics. A classic example would be a record group of Enquiry Commissions appointed by the government. Each such commission is a unit comprising the whole of its own organic records, but for archival administration it is more efficient and much easier to group all such commissions together under one descriptive roof. This will be an artificial record group composed of organic cells.

The last and fourth type of record group would be the documentary record group, which in reality is a collection of different documents, files and material of different or special formats. A documentary record group is subject or format oriented, bringing together in a non-organic way documents on a relevant topic, territory or personality, or of a special format, such as photographs or films.

It should be stressed, however, that personal papers can be an organic record cell - though very often unorganized

- and in most archival institutions they are grouped as organic cells - sometimes with collected additions - in artificially composed record groups. In libraries, however, personal papers are very often handled as collections of manuscripts without any intention of preserving organic growth.

I would like to add a few words about the sometimes heard thesis that the record group concept is not valid any longer and one should arrange records according to series. If the meaning of the word "series" is one record-keeping system, one record cell, then I agree on the theoretical basis, but I have my doubts about the managerial aspect of our discussion. Instead of a few hundred record groups listed in a guide and forming the basis for archival management, we would have to deal with many thousands of series. But more serious is the use of the term "series" as a descriptive unit when a number of series are in reality one organic cell and this proposed approach of putting the series in the place of the record group would make impossible any archival description and logic arrangement.

Twenty-five years ago Oliver Holmes published a classical paper on Arrangement of Archives⁵, though some of his conclusions are no longer valid. The Israeli practice developed, at the time he published his paper, a more pragmatic approach based on the same principles. In our archives the depository level means "storing area" and by now we have at the State Archives 15 such areas which serve managerial purposes only. All court records, for example, are in one store unit. Sometimes one record group can be divided physically between two or even three storing areas. For example, files of the Prime Minister's Office are partly in a special storing area of classified security material and some of its audio-visual records in the storing area reserved for these formats. The location symbol of individual items will always be composed of the symbol of the storing area and the container (box) number within this area. Within one storing area there are many record groups which form units for administrative and descriptive purposes. A record group is described by information on its administrative history and the main functions of the administration. The main descriptive units - on the other hand - are the series, sub-series and files.

As said before, there is no connection between physical arrangement on the shelves, which is reflected in the storing units, and their numerical lists only, and the description which is based mainly on series and related to the record cells and the record group as a whole.

In concluding my remarks I wish to say a few words on the problem of transfer of functions and record series from one organism to another. On principle the last administrative organism in which the records were actively accumulated is to be considered the originating agency of a series. Cross references from one record group to another will always solve the problem. With the help of computerized lists and inventories this is certainly no problem. Functions which tend to migrate - together with their records - from one organism to

another are best considered record groups in their own right, even if at times they did not have the characteristics of an organic administration of their own. □

¹ARCHIVUM vol. XIV (1964) p.16.

²Scott, Peter J. The Record Group Concept: A Case for Abandonment. *American Archivist*, vol. 29, Apr. 1966.

³Brenneke/Leesh. *Archivkunde*. Leipzig, 1953, p.83.

⁴Fenyo, Mario D. The Record Group Concept: A Critique. *American Archivist*, vol. 27, Jan. 1964.

⁵Holmes, Oliver W. Archival Arrangement - Five Different Operations at Five Different Levels. *American Archivist*, vol. 27, Jan. 1964.

National Archives and Electronic Records: Where Are We Going?

by Sue Gavrel¹

National Archives of Canada

Introduction

The information society has had a major impact on the activities of traditional archives. The prediction of the "paperless office" in the eighties has not materialized as yet and in fact the amount of paper has increased substantially over the past decade. Rather than reduce paper, the introduction of computer technology has increased the number of products and copies of those products. The work of the archivist in the identification of the archivally valuable records has increased due to the paper burden. One must sift through far more records to identify those of historical value.

Added to the increase in the number of records created is the pressure of the research community to retain more rather than fewer records. Prior to the mid-seventies, the major factor in the appraisal of records was that of evidential value - the evidence the records contain of the organization and functions of agencies. Archives (and I restrict my comments mostly to North American and particularly Canadian Archives) have acquired many more records based on their informational and research value over the past fifteen years. In Canada, this coincides with the growth of social and economic programs of the federal government.

The computerization of many government programs began in the early sixties and has increased ever since. The centralization of edp expertise was very evident during this time and continued until the arrival of the micro computer. The large database systems were, in most cases, built by the EDP experts and used to service the program managers' needs.

Archival Programs for Machine Readable Records

The use and importance of computers was recognized by the large National Archival repositories in many countries in the establishment of machine readable record programs. Over the years standards were developed for the appraisal, acquisition, processing, conservation and servicing of machine readable records. Due to the small number of archivists involved in these programs, a great deal of co-operation and sharing of information lead to the development of procedures to handle these new records.

Only a minimum amount of success was achieved in the identification and subsequent transfer of computer records of archival value from government agencies. The control of these records was in the hands of the edp area and outside the normal channels of the control of paper records (Records Managers).

Efforts to mimic the systems in place for the control of paper records met with limited success, mostly due to the lack of familiarity with the archives by those in charge of the development of the systems. Machine readable programs, in traditional archival settings, although recognized as important, lacked the focus and strength required to affect the overall organization of records.

Trends

In recent years, this trend is changing due to a variety of reasons. In the discussion which follows, I will outline some of the changes and trends which will have a profound impact on archival repositories. These changes range from new understandings of the importance of information; technological change; and major changes in the ways data are created, stored and used. The next decade will require archives to focus on electronic records or risk losing the electronic cultural heritage.

Technological Change

It is not my intention to provide an overview or history of the changes we have experienced in technology in the past decades. It is, however, important to review some of these changes in light of how records are created, why, and how archives must adapt to these changes. The major trend which has affected the way records are created results from the rapid penetration of microcomputers into the market in the last five years and, in particular, into government departments and agencies. The centralization of edp services is disappearing with the use of microcomputers in the office environment. Managers, officers and clerks have now as much computing power on their desks as the mainframes of the seventies provided. The ability to create, manipulate, access and disseminate data has been decentralized. The trend to purchase "off the shelf" software has taken away from the centralized edp shops in the creation of in-house software and database systems. Linked to the penetration of microcomputers is the development of local area

networks. The linking of staff provides for the creation and revision of documents which can be done on-line with only the final version being available in either hard copy or electronic format. The ability to provide for the development of documents relating the evolution of the policy, to changes in administration, or to data is now in the hands of the creators of those documents. In most cases, LAN's do not provide for the traditional records management approach to control of records.

Individual workers make decisions regarding the disposition of the records. Such a system existed for the development of large database systems under the control of edp professionals. The difficulties experienced in gaining control over what information is being created and destroyed can be magnified as all employees become responsible for their records. The program of economic restraint experienced in most western countries is also leading to the use of microcomputers, as it is seen as a way of increasing productivity and decreasing personnel costs.

A major effort can be seen in the development and interest in communication standards. The lack of compatibility between hardware has been, and continues to be, a major problem to the increased usage of data. The efforts now seen in the development of International Standards is encouraging. The trend to the Open Systems Interconnection standard protocols provide for the possibility of connecting systems with different hardware. Other standards will have an impact of the increased sharing of data. Map and Chart Data Interchange Format, or MACDIF, data is an attempt to provide a standard format for the transfer of chart and graph data from and to a variety of systems. Office Document Architecture/Office Document Interchange Format (ODA/ODIF) provides for similar transferability of text. The work towards developing and implementing such standards must be followed closely by archivists, as it is through such efforts that some of the technical issues such as making valuable data accessible in the future may be resolved.

New techniques for software development such as fourth generation languages and expert systems techniques are becoming important tools providing faster and more flexible software development and more user interfaces. No longer must the design and development of databases be the sole responsibility of the edp professional. Databases can be created and used by those who have access to D-Base or other such software. Expert systems potentially pose major problems for the archivist. In the past the acquisition of data tried to steer away from software dependent systems. Expert systems which can be defined as "an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution" are only in the early stages of practical application. Their potential to assist managers with complex planning and scheduling tasks, diagnose diseases, etc. is great. The impact of such systems on the documentation of the decision making process is evident. How archivists will respond to such systems is a major

challenge.

Types of Data

Part of the technological changes, but one which should be highlighted, is the trend towards integrated systems and applications. Two specific types will be discussed:

The Geographic Information System (GIS)

Geographic Information Systems are beginning to play a major role in the information society. Today's systems only superficially resemble the automated mapping systems of the sixties. GIS are increasingly being used to conserve and manage a wide variety of data from natural resources to environmental pollution as well as in the planning and management of cities - such land information systems cross organizational and sectorial boundaries and represent an opportunity to develop new information based products and services.

Compound Documents

The move to more integrated systems is seen in the "compound document". The integration of voice, data documents and graphics oversteps the traditional media boundaries. All are reduced to the common language of binary code. Not only is it feasible to create the compound document, but it may also have been created from information which was only accessible on the screen for a brief period. The source of that information cannot be traced. The accessibility of data from other systems through local area networks and the merging of data from a variety of data bases will create documentation problems for the archivist. The ease with which such information becomes available and usable will be reflected in the move to adopt communication standards, more user oriented software, and more computing power.

Information As A Resource

The information society has created a new awareness of information as resource. The major expenditures on hardware and software development of the seventies has created a new awareness of the value of the information which these systems manipulate and store. In Canada, Access to Information and Privacy legislation led to the acceptance of a computer based record as a record. In the definition of a record for the purpose of the legislation, machine readable is included.

The requirement to account for information regardless of the machine on which it was stored was an important step in the recognition of computer records. The new National Archives Act passed in 1987 also uses the same definition of record. The Act stipulates that no records of the Government of Canada can be destroyed without the consent of the National Archivist. It further stipulates that those records deemed to have archival value must be transferred to the National Archives. The definition of record to include machine readable records ensures that electronic records are part of the National Archives' responsibility. More recently two new policies have

added to the importance given to information: the Information Holdups Policy (just recently approved) which will direct government departments to manage their information in a holistic manner and account for that information through the development of directories to it; and the Information Management Plan which outlines to departments the importance of the planning of information management rather than the justification of new equipment purchases.

All of these policies, acts, and planning strategies are the result of an awakening to the importance of information as a resource.

Another interesting trend in the field of information is the move by the public and private sectors to develop co-operative databases. The Geographic Information Systems are an example of this type of co-operative effort. Similar to the compound document such co-operative databases will have an impact on the archival organization of records.

Much more could be said on the trends and changes which technology will initiate. It is important to look briefly at the impact such changes will have on the traditional functions of archives.

The Challenge to Archives

"The digitization of information through the common language of the binary code is bringing about the convergence of voice, usage and data - and of the telecommunications, electronics and computing industries based upon them"².

Over the last decade, Archives have tended to expand according to media-based responsibilities: textual records, cartographic, film and television, photographic and machine readable. Practices and procedures were developed to acquire, process, store and service the different forms of information as each had its special requirements. The fundamentals of archival theory were common to all media. Appraisal criteria, were based on the principles of evidential, informational and legal value. It was in practices for arrangement and description where the differences became more evident. Procedures and practices for the long term preservation of machine readable records were developed in the seventies. The procedures were based on large mainframe systems and proved successful for the conservation of data in systems.

Technology is now the driving force behind the integration of the different types of records. Just as media divided Archives into specific units, it now will play a large role in integrating these units. With the use of electronic technology in the creation of all types of records, the media on which the information resides becomes the common element.

Information is created and transmitted in so many forms that archivists must now look at program activities as a whole and identify those records which have archival value as well as the most appropriate form in which they should be stored. Electronic records provide many

research possibilities. As more types of records are created in electronic form - more such records are likely to be of archival value. The major obstacle is, of course, the long term accessibility of the information in electronic form. For other records-paper, photos, maps established techniques have been developed which preserve the records for future use. Technology did not have a major effect on long term preservation except for improving the techniques. With electronic records, the media, the software and hardware are constantly changing - Evidence of this can be seen in the experience gained to date. Electronic records created in the seventies are different from what is now being created. Procedures, valid for data in systems, must be modified and reevaluated to cope with compound documents and GIS. The technological requirements put pressure on established procedures. The focus of any archival program for electronic records must focus its resources on resolving the technical issues of how best to transfer records to the archives, how to process these records to ensure their accessibility. Archivists will be required to support efforts to ensure standards; to become involved with systems as they are being created; and to keep abreast and knowledgeable about the changing technology and how it affects the creation and use of records.

These are major changes for institutions which have traditionally dealt with the past.

Finally, new methods in records creation may have fundamental effects on traditional archival theory and principles. Archivists must become active participants in the creation of information, in many instances identifying elements of archival value before they are created, in order to ensure the preservation of the historical record. Archives have, to date, been concerned with documenting the activities of an organization or business. How does this new role affect the documentation of activities when the archivist has participated in the creation stage?

As more organizations undertake co-operative efforts in information creation, how do we determine which records originate with which organization? Who has the ultimate responsibility or control of the records? Such systems break down the barriers between public and private sectors; federal, provincial and municipal governments. The clear lines of origin become blurred.

Conclusion

Today's presentation can only briefly mention the issues and resulting challenges to traditional archives. Issues such as these are of utmost importance to the archival community. The "paperless office" is not yet a reality but signs of its existence are much more evident today than they were two to five years ago. Efforts to resolve the problems are imperative if records documenting the nineties are to be available for future generations. □

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989

²Forester, Tom. High-Tech Society. MIT Press, 1988. p.1

Data in the Natural and Exact Sciences and in the Social Sciences; a Comparative Study

by Eliahu Hoffmann¹

National Center of Scientific and Technological
Information, Tel-Aviv, Israel

Introduction

A comparison of data in the social sciences with those in the natural and exact sciences seems at first completely out of order. The nature of the data in both areas of science is so different that, content wise, they certainly cannot be compared. However, if we approach the comparison from a technical operational point of view we may try and classify them and compare.

1. the types of data that we find in both areas of science,
2. the treatment that we accord these data in both areas of science.

Such a comparison may be possible and may furthermore yield deepened insight and understanding in the meaning of the data. In addition, operational technical experience in data handling may be exchanged eventually between the two areas of science and this in turn may improve methods of data handling in both of them.

Types of data

In the natural and exact sciences three basic types of data can be discerned.

1. Fundamental constants and permanent data,
2. Time and space (or place) dependent data,
3. Factual data.

1. Fundamental constants are constants which have been measured directly or indirectly presumably "once and for all". They appear in numerous calculations and therefore their precise value is of great importance. If their values have been established a long time ago it may happen that, with better measuring instruments and methods they are being slightly corrected or improved. These corrections usually touch some of the last digits and thus make them more precise. By and large though, their values do not change anymore. A few examples may be quoted such as: The speed of light (299,792.8 km/sec or 300,000 km/sec)²; Planck's constant (h) which represents the elementary quantum of action in physics³; and even the

standard length of the meter (0.0000001 of the earth's meridian quadrant at sea level)⁴.

Apart from these, there exist numerous series of permanent constants which have been measured and verified and are valid ever since. As examples we have: The values of the atomic weights of the elements; the infra-red (I.R.) absorption spectra of chemical compounds. Each compound, according to the structure, exhibits a series of absorption lines of specific wave lengths of the infra-red spectrum. These are determined by the structural characteristics of the compounds and serve as analytical tools to identify the compounds and their structures. Similar data series are: the solubility data i.e. the maximum quantity of material soluble in a certain volume of liquid. This quantity usually rises with the rise of temperature of the liquid and is constant for every temperature of the liquid.

2. Time and space (or place) dependent data are only true and valid for the place and time at which they were measured and describe a certain numerical situation at that place and time. Intensity measurements of incident sunlight may be taken as an example. These are usually high in summer and low in winter, higher in the equatorial regions and lower in the northern and southern regions of the earth and strongest at noon and zero at night. In addition, ecological data belong to this type i.e. the concentration of carbon dioxide in the air is high in urban industrial areas and low in rural areas. Data on rainfall are of a similar nature i.e. dependent on the time and the place when and where it occurs.

3. Factual data are essentially descriptions of properties of substances and/or processes. Among them can be named the properties of materials, whether they are hard or brittle, elastic or rigid, whether they are heat and/or electricity conductors or isolators etc. The taxonomic categories of classification of living organism belong to this type as well as elaborated DNA sequences in genetic material.

It seems that in the social sciences similar types of data can be recognized.

1. Constant data here too are data the validity of which is permanent and not dependent on - or restricted to - a specific time and place. (Though the data themselves may represent either dates or places). Personal records

for instance may be cited as an example. Thus social security numbers, identity card numbers, dates of birth, dates of immigration/emigration; educational records such as dates of awards of degrees etc. are all essentially constant data.

2. Most of the data in the social sciences however seem to be of the time and space dependent type. Census data, election results, manpower surveys, import-export statistics, fiscal and financial data, all these relate specifically to a certain time or time - period and to a certain place - city, country, continent etc.

3. As for factual data in the social sciences; historical and archival records, laws and regulations etc. can be considered to fall into this category.

Thus in spite of the fact that the data in the social sciences are very different in character from those of the natural and exact sciences, the same type of classification and characterization can apparently be applied to both.

Treatment of data

In the natural and exact sciences this consists essentially of 5 distinct steps:

A.) collection, B.) evaluation, C.) organization, D.) storage, E.) retrieval.

A.) Collection: Data in the natural and exact sciences are usually "buried" in textual research papers and have to be traced and pulled out. Thus, properties of materials or spectroscopic data, for instance, even for one substance only, may be found in a number of different papers since they are recorded also for analysis and identification purposes. If a list of data on one specific property for a large number of compounds is to be prepared, the search for them will extend through an even larger collection of research publications.

In the social sciences, by comparison, data are usually either collected and recorded routinely for well defined purposes. For instance, car license and ownership are registered, *inter alia*, for statistics of vehicle distribution; export - import data are used for balance of trade calculations; issue of identity cards serve population statistics and passport control shows the movement of people. In addition, they can also be collected by means of special projects i.e. surveys etc. In both cases the relevant data are obtained, *a priori*, in a more centralized fashion.

B) Evaluation. Since data in the natural and exact sciences are collected from a variety of sources, their evaluation is of great importance. From a systemic point of view one can recognize here the following 4 aspects to be checked.

B1) Conditions of measurement: These include the type of instrument used and the calibration of this equipment, the type of experiments performed and the methods by which they were measured. Were the correct, the relevant parameters measured? Were standardized

measuring units used? Are the results reproducible? Have the experiments been repeated and how often?

B2) Methods of calculations: What kind of approximations have been used and what margins of error have been specified? Are both of them compatible with the conditions of the experiment and the precision of the measurement?

B3) Methods of presentation: Have the results been presented in a uniform fashion? What scales have been used, logarithmic, arithmetic? Can the data be read precisely from the graphs or tables used?

B4) Reasonability of the results: These should be judged in comparison with similar experiments and measurements. In chemistry, for instance, some physical data depend on the structure of the compounds. Some specific structural groups exhibit distinct and unique absorption in the infra-red spectrum and the appearance of such an absorption is taken as proof that the compound has indeed this structure.

Evaluation of data in the social sciences is apparently somewhat different. Conditions of measurement (B1) are presumably fixed in advance, either by routinely collecting specific parameters, like identity card numbers, or by specifying the parameters to be collected via surveys, by means of interviews and questionnaires. Recently, however, it has been recognized that surveys carry with them possibilities of error due to the "cognitive processes that respondents are required to exercise" i.e. to interpret the question. These errors may be more prevalent with questions for information than the queries for numbers, but nevertheless they should probably not be overlooked. Methods of calculation (B2) like approximations and margins of error can also be specified and/or decided upon in advance and depend *inter alia* on the sample size and the response to the questionnaire. Methods of presentation (B3) can be chosen freely so that the essence of the results and their interpretation are emphasized and finally the reasonability of results (B4) depends *inter alia* on the tracing and elimination of technical and systematic errors.

C.) Organization: The organization of data is crucial for their efficient retrieval and proper utilization and in general can provide insight into their significance. Listings of data in the natural and exact sciences are usually presented in tables which are collated in Handbooks. In chemistry, for instance, properties of compounds - like boiling points or melting points - can be arranged in tabular form according to the names of the compounds and in alphabetic order or, according to their molecular weight; in this case from the lightest to the heaviest compound. This order would of course be different from the previous one. In most cases, the alphabetic order according to the names is sufficient, preferred and the only one used. A number of physical constants - and not only one - are usually tabulated alongside the name of the compound. Updating, however, must be provided for the addition of new compounds and their properties. The organization of data in the social sciences seems to be more complex and

multifaceted. In the natural and exact sciences orders and arrangements according to a single parameter, normally a name, are quite sufficient. In the social sciences, however, the ordering and arrangements according to specific groupings are very important because it is by means of these rearrangements that the insight, the meaning of the data become evident. Population data, for instance, must ultimately be arranged according to: a) Age groups; b) Communities; c) Areas of population; d) Professional groups; e) Groups of income etc. if one wants to understand their significance. On the other hand, there is here no need for updating because the data describe a certain situation in space and time which cannot be repeated at will. New updated data shall describe a different - time and space - situation.

D. Storage: Data in the natural and exact sciences are stored in printed form in Handbooks. They have to be characterized by specific terms, key words, identifiers etc. for tracing and retrieval by means of indexes. The same holds true for their storage in computerized form; a number has to have a computer address, a flag, an identifier etc. Basically these requirements for storage are identical in both areas of science, simply because they are a function of the operational methods of information storage (and retrieval) and are essentially independent of the subject matter at hand.

E. Retrieval: Methods of retrieval of data too are identical for both areas of science. For the printed form the indexes are used and for the computerized database the retrieval and search programs perform the task. They are strictly methods of information processing, independent of subject matter or content.

Similarities and differences

In the previous paragraphs it was shown that data in the two areas of science can indeed be compared. The types of data are categorized in identical fashion and their treatment exhibits considerable similarities. These results would seem surprising in view of the basic differences in the subject matter of the two areas. One is indeed tempted to ask: "If they are so similar, how do they differ?"

It appears that the difference lies in the approach to, and the emphasis on data which prevail in the natural and exact sciences. This in turn may be due to the following two factors:

1. The natural and exact sciences have more permanently valid data than the social sciences and that from the early beginning of their development. Indeed, measurements carried out in chemical experiments one hundred years ago are usually valid today and may be repeated with the same results. The time and place dependent data which constitute the majority of the data in the social sciences are of more recent origin in the natural and exact sciences, to mention the environmental and ecological data as an example.

2. The quantities of the available permanent data are simply "mind boggling"! In chemistry, alone, for instance, at least 6.5 million distinct compounds are known today. Assume that each compound has at least 10 single constants and 1-2 series of permanent data (infra-red data, solubility data etc.) and you have hundreds of millions of numerical data.

For this reason the proper treatment of data in the natural and exact sciences, though not an aim in itself, was always considered of paramount importance. So much so that within the International Council of Scientific Unions, the umbrella organization for all the scientific international societies, a special scientific international, interdisciplinary organization was established. This organization, called "The Committee for Data in Science and Technology" - CODATA, is concerned with the theory and practice of the proper methods for data handling. □

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989

²Parker, Sybil F., Editor in Chief, McGraw-Hill Encyclopedia of Science and Technology. 5th ed., McGraw-Hill Book Co., New York, N.Y., 7:683, 1982.

³*ibid.* 10:368

⁴*ibid.* 8:486

⁵Fienberg, Stephen F., Tanur, Judith M., "Combining Cognitive and Statistical Approaches to Survey Design", Science 243 (4894):1017-22, 24 February, 1989.

Policy-Oriented Research in Britain Some Institutional Mechanisms

by Robert Barnes¹
Office of Population Censuses and Surveys
London, England

Introduction

In Britain initiatives for policy-orientated survey research come from the the major departments of state with social responsibility - for example Departments of Employment, Education, Health and Social Security, and the Home Office. The research itself is sometimes carried out within government agencies, sometimes it is contracted to outside agencies and sometimes it is a collaborative effort. There are no completely consistent criteria which determine where the research should be undertaken - the decision usually depends upon the nature of the research, where the expertise lies, the nature of the policy which gives rise to the research, the availability or resources within and outside government and what trade-offs the commissioning department wishes to make between speed, cost and quality.

As in many countries of the world, quantitative social research for policy purposes in Britain relies upon a number of data sources - that collected by surveys, by population and housing censuses, administrative records and so on. Because of the different strengths and weaknesses of each data source, they are used in different ways in the social policy framework. For example data needed for *immediate* input to policy invariably need to come 'of the shelf' and so tend to come from existing administrative sources, census material or from multi-purpose household surveys - and most of these data and the research tend to be in government. But data needed for longer term, strategic policy initiatives - or for policy monitoring or evaluation purposes - may often depend upon more directly 'tailor made' research work and for this the research may be undertaken within government or it may be commissioned from outside government. Where data are needed quickly and do not already exist, it is sometimes possible to meet the need to a limited extent by buying questions on one of the 'omnibus' surveys run by commercial research agencies.

Mechanisms outside government

Outside government there are three main alternatives that might be used by policy departments to undertake survey research: academic researchers within universities, specialised research institutes, and market research

companies. Generally the universities and research institutes have limited or no operational facilities to mount large scale data collection exercises. Where the research requires this they will need to work collaboratively with government or market research agencies which do maintain these capabilities.

For example some years ago the British Treasury contracted an academic economist from a British university to carry out research into the incentive/disincentive effects of direct taxation on labour supply. The research required high quality data for the development of an econometric model and to obtain such data the researcher approached the specialised government survey organisation, Social Survey Division (SSD) of OPCS, to collect and process the data and to produce machine readable tapes. This was a successful collaborative exercise in which the divisions of labour were clear: the research was designed and analysed by the academic expert while the technical operations were undertaken by the specialised organisation within government.

In recent years the government statistical service as a whole and the statistical activities of OPCS have been subject to a number of reviews by scrutineers appointed by government. One of the recommendations has been that there should be moves to put more survey research out to contract with the private sector. The latest estimates available are that probably over 90 percent of all government surveys are done by outside agencies although because most of the large surveys are still carried out by SSD, the amount contracted out, by cost, is about 50%. Of the amount contracted, over half is done by market research companies. The rest is commissioned from academic researchers, private consultants, research institutes and the like.

The links between university based social science and practical based survey research, are probably less developed in Britain than in a number of other countries. However there have been a number of initiatives intended to reduce the gap. For example there is now an 'academic link' scheme whereby facilities are made available to enable academic social scientists to spend some time observing operations in specialist survey research organisations to gain direct experience of the practicalities of carrying out surveys. Also the Economic and Social Research Council maintains a data archive, located at the University of Essex, in which are lodged

nearly all the main social data sets including population censuses and surveys. The Archive makes these data available for secondary analysis purposes by academic researchers.

Much more recently plans have been announced to establish, also at the University of Essex, an Interdisciplinary Research Centre (IRC) to bring together teams of social researchers working on a variety of projects connected with the measurement and analysis of social change. An important element of the IRC work will be to set up a new panel study of some 5000 households a year who will be re-visited over a number of years to produce new sets of data for longitudinal analysis on a variety of topics - many of which will be of direct relevance to policy interest of government. The plan is to launch the new survey in 1990 and it may be in time that the IRC will develop its own operational capabilities.

Mechanisms within government

Within government there are research units within policy departments and there are specialist government agencies so it is helpful first to look at the way government statistics and research are organised. In Britain government statistics are largely decentralised so that each policy department has its own statistical division and these collectively, and with the Central Statistical Office, the Business Statistics Office, and the Office of Population Censuses and Surveys make up the government statistical service. In addition a number of these departments also have their own research units such as the Home Office Planning and Research Unit, expert in criminological and penal research, and the research units in departments such as Health, Social Security and Employment. These research units are responsible for initiating and carrying out policy related research for the departments in which they are sited, but they do not possess operational research facilities for data collection and processing activities. For that they need either to commission one of the commercial research agencies or the Social Survey Division of OPCS. Whichever option is chosen the departments, through the research or statistical divisions, (and sometimes through the policy/administrative divisions) may either commission the whole research job including design, analysis and preparation of interpretative reports, or they may contract out just those technical activities such as data collection, that they do not have facilities to do for themselves.

Frequently departments will call up SSD for advice and assistance in handling the contracting arrangements. This advice may cover various stages of the process including consideration of which organisations to invite to submit tenders, drawing up research specifications, interviewing tendering organisations about their bids, evaluating tenders and, sometimes, monitoring the performance of the successful organisation in carrying out the work.

Social Survey Division of OPCS

SSD is the government's own specialised survey research organisation. It began in 1941 as a separate organisation within government for carrying out survey research work. In 1970 it was amalgamated with what was then the General Register Office to form OPCS and it is now a division of that Office.

Nowadays most major government household surveys are undertaken by SSD and for most projects SSD is responsible for all stages of the survey enquiry, including the original definition and clarification of the survey objectives, project design, sample selection, data collection, processing and finally the production of a report on survey results. The full cost of work carried out has to be paid for by departments whether it is commissioned from SSD or contracted out. It is frequently the case that contracts are subject to competitive tender and it is now possible for SSD to submit tenders and so compete directly with private agencies.

SSD carries out surveys on a wide range of subjects for government departments which have responsibilities for social policy. Most of these surveys are of a point in time, or *ad hoc*, nature, but some are longitudinal, visiting the same sample over several points of time, and others are conducted on a continuous basis throughout the year and every year. The division is responsible for 6 major continuous surveys.

- a. The Family Expenditure Survey (FES) has run continuously since 1957 and seeks interviews at just over 11,000 private households in Great Britain a year. The main content of the survey is household income and expenditure and the chief purposes are to provide up-to-date expenditure weights for the Retail Prices Index and to enable government departments to analyse the economic and social effects of policy changes in taxation and social security benefits.
- b. The National Food Survey (NFS) has run since 1940 and currently seeks interviews with around 15,000 housewives a year. The main content of the survey is the purchase of food and the data are used mainly to calculate consumer expenditure on food for national accounts purposes and for economic and nutrition analysis.
- c. The General Household Survey (GHS) has run since late 1970 and has a set sample of around 12,500 private households in Great Britain a year. It is truly a multi-purpose undertaking and has no dominant subject theme. The main areas covered are demography, housing, employment, education, health and income but there is too a great variety of subsidiary topics which are included from time to time such as leisure activities, dental health, burglary, smoking and drinking behaviour.
- d. The International Passenger Survey (IPS) began in 1961 and is a multi-purpose enquiry carried out at all major airports and seaports in Britain. It collects information from an achieved sample of about 165-180,000 travellers each year. Data from the survey are used for a

wide range of statistics on tourism, for estimating the tourism component of the balance of payments and for statistics on international migration.

c. The Labour Force Survey (LFS) started in Britain in 1973 as part of an already established series carried out biennially for the European community. From 1983 it has become an annual enquiry. The sample for the survey comprises some 15,000 households a quarter with a 'boost' of 45,000 extra households during March to May to give a total sample of 60,000 for that quarter.

f. The National Travel Survey (NTS) is a continuous travel survey that started in the summer of 1988. The set sample is around 5,000 households each year and the survey is to provide a continually up-dated set of data about personal travel in Great Britain.

Altogether the continuous surveys account for about three-quarters of the total SSD budget. However the *ad hoc* side of SSDs activities cover far more projects in number although the sample sizes are nearly always much smaller. It is possible to distinguish several kinds of *ad hoc* projects.

1. There are those which deal with the physical condition of the population. For example a major survey programme was carried out between 1985 and 1988 to provide estimates of the prevalence and conditions of disabled adults and children. The results will be used to help determine policy on state benefits payable to the disabled. Other examples of surveys under this heading are those on dental health, drinking, smoking, nutrition, and the heights and weights of the adult population. Results from these surveys are used for a variety of policy purposes including health education and prevention medicine.

2. There are those which deal with social circumstances. Recent examples here have included the effects of unemployment on family living standards, the financial (and other) consequences of divorce, and the circumstances of people who have recently moved into privately rented accommodation.

3. Thirdly there are surveys designed to seek out people's opinions and attitudes. Examples here include surveys to understand better why children smoke, the attitudes of women to maternity services, and to provide estimates of the demand for higher education.

4. Finally there is a miscellaneous group of *ad hoc* surveys. This includes job content studies on particular groups such as junior hospital doctors and community nurses, enquiries of visitors to public facilities such as museums and galleries to help management plan and improve the services provided, and census related studies such as question wording tests and post enumeration studies. In addition there are studies which monitor administrative practices in local offices to provide information on the extent to which central policy or guidelines are being followed locally: examples of this are enquiries into the way electoral registers are compiled

Surveys carried out by SSD are characterised by emphasis on minimising non sampling errors² For example samples are nearly always random, self weighting probability samples and for samples of the general population the frame used is the Postcode Address File (PAF) - a virtually complete listing of every address in the country, available centrally in machine readable forms. In fieldwork emphasis is placed on using the best calibre interviewers (approximately 6-7% of applicants pass through all stages of recruitment and training to become part of the regular fieldforce) and training. All surveys are voluntary and considerable attention is paid to achieving the highest response rates possible. It is also standard practice to carry out pre-tests and pilot studies of questionnaires and proposed overall survey designs before embarking on the main stage of surveys.

Work concerned with the development and evaluation of government campaigns of social persuasion and communication (on subjects such as drinking and driving, drugs, safety, and recruitment to the uniformed services) is not done by SSD. This is carried out by the Central Office of Information (COI) which is the department responsible for government publicity and advice on publicity. The COI does not have its own fieldwork resource and so is a major contractor of work to the private sector.

The population census

The population census in Britain is carried out on compulsory basis every ten years. The census is conducted by OPCS using conventional drop-off/pick-up methods with over 100,000 temporarily recruited enumerators. There is no field sampling but data processing is on a 10% only basis for the most difficult to code items such as occupations.

Results are made available mainly either in the form of pre-determined statistical abstracts (published or unpublished) or in the form of specially ordered tabulations. So far in Britain there is no access to individual records although discussions are currently in progress about the possibility of making available samples of anonymised records for the next census in 1991, and for establishing on-line access to the census data base such that tabulation parameters can be fed in and tables obtained.

It is possible to use the census as a sample frame for identifying particular groups of the population for more detailed enquiries - for example particular occupation groups. The fact that the Social Survey and Census are two divisions of the same Office makes follow-up enquiries simpler to do without arousing public anxiety about infringing the confidentiality undertakings given at the time of the census.□

¹Presented at the IFDO/ASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.

²Barnes - Non sampling errors: some approaches adopted in major government surveys in Britain. *Journal of Official Statistics*, 1987.4. Statistics Sweden.²

U.S. Bureau of the Census Data Dissemination Programs

by William P. Butz¹, Associate Director
for Demographic Programs and
James R. Wetzel², Chief, Center for Demographic
Studies U.S. Bureau of the Census

U.S. Census Bureau Data Dissemination Programs

I am pleased to join you today to discuss the data dissemination and public data user support practices of the U.S. Bureau of the Census. Our policy in this area is straightforward and very simple. We strive to provide maximum public access to and use of data we collect, while assuring the confidentiality of respondents and satisfying budget constraints. We maintain an extensive data user support organization, provide user training, and make data available in many forms-printed, microfiche, and all electronic media.

The Bureau staff is inventive and aggressive in its efforts to maximize data access. Among other firsts, we were the first statistical agency to release public-use microdata files from household surveys and censuses (in the early 1960's) and most recently, the first U.S. government agency to release files on CD-ROM. Today, I will describe our basic data dissemination programs and share with you some thoughts on our ongoing efforts to extend the use of our microdata archives by providing analytic access to longitudinal research files on the characteristics of manufacturing firms and microdata from household surveys and censuses for smaller geographic areas.

The Right Data in the Right Form

Our data dissemination policies are expected to simultaneously fulfill a broad spectrum of highly challenging objectives including:

We are expected to provide low-cost, accurate data;

Our data should be in an easy-to-use form;

We are expected to ensure timely delivery;

We need to provide technical support to users ranging from elementary school students to highly sophisticated, computer-based reproducers of sample data; and

Our systems must work for a broad range of data and must be flexible to deal with large fluctuations in our output.

To meet the needs of our broad spectrum of users, data are made available in six separate forms: printed reports, summary tape files, CD-ROM, microfiche, public-use microdata tape files, and customized special tabulations. All forms for output are rigorously reviewed and edited to ensure confidentiality of individual data.

Printed Reports. Printed reports are the dominant medium for the bulk distribution of statistics to users. They reach a broader audience than data in machine readable form, and our society still is better equipped to handle - by way of libraries, bookstores, and photocopying machines - the dissemination and maintenance of data in print. We publish between two and four thousand reports a year (ranging in length from one to 1,500 pages). We frequently come to the limit of what we can afford to put into print long before user demands for decision-relevant statistics have been fully satisfied. This problem, which derives in part from high printing and distribution costs, is exacerbated by financial disincentives built into federal printing policy. In brief, when we publish a report, we must provide from appropriated funds all of the capital costs of the publication and make copies available at no charge to various federal agencies (depository libraries, for example). We do not receive a penny from subsequent Government Printing Office sales of the reports. Thus, the more statistical reports we publish, the more it costs us, regardless of how well we may be meeting a need. And we clearly are meeting needs; otherwise the government would not sell 35,000 or more copies of publications such as the Statistical Abstract annually.

Summary Tape Files. Summary tape files (the second major medium for data dissemination) actually are a by-product of processing data for printed reports. These machine-readable data files contain frequency counts or aggregates similar to those in printed reports. In many cases, these files provides much more statistical detail than is available in final printed reports; the files often are more timely than printed reports; and, of course, they already are in a form conducive to automated analysis. The costs of data collection, and of designing, producing and thoroughly documenting the files are part of functions supported through our regular funding. Since there are no printing costs and few compulsory distributions, the disincentives which make printed reports so expensive for us do not apply to machine-readable data files.

To maximize the utility of data in this cost-effective form, we distribute selected, heavily-used files to fifty-four State Data Centers (including District of Columbia, Puerto Rico, Guam, and the Virgin Islands) to help make this resource widely available to users without computer facilities.

CD-ROM. Digital laser-disk technology is just beginning to be used by the federal government for dissemination of large files. CD-ROM (compact disk read-only memory) readers are now on the market that allow a microcomputer user to search and read data bases from a single disk holding over 500 million characters, equivalent to the combined storage capacity of at least three high-density tapes. Further, since these disks allow "random access," any piece of information in an appropriately indexed data base can be retrieved immediately, unlike data on "sequential access" tape, where every preceding piece of information must be read first before the desired data can be accessed. The price of CD-ROM laser-disk readers is likely to decrease further, but already they are well within the reach of business microcomputer owners.

Microfiche. The needs of data users with bulk holdings, such as libraries, frequently are met through a third medium, microfiche. Microfiche are much less expensive than printed products; they display much larger quantities of statistics in a given amount of storage space; and retrieval of selected statistics can be accomplished without the technical sophistication required of a computer user. Most data from the 1987 Economic and Agricultural Censuses and the 1990 Decennial Census will be available on microfiche.

Special Tabulations. While the quantities of statistics available in print, microfiche, computer tapes, and other media have increased steadily for many years, there are inevitably specialized needs we cannot hope to anticipate with general-purpose products. To deal with such situations, the Bureau maintains a capacity to respond to user requests for customized statistical output, assuming, of course, that the user is prepared to reimburse the Bureau for production costs. I am not talking about the kinds of statistical compilations that could be performed by someone else using products in the public domain, but rather those tabulations which require use of the Bureau's confidential basic microdata on establishments, persons, or housing units.

All special tabulations are treated as an extension of our publication program by eventually making copies available at reproduction cost to anyone after delivery to the sponsor. Thus, they sometimes provide a way for us to prepare data products which we would like to have available to users, but which cannot be prepared within appropriated funds. For example, ZIP code tabulations from the 1980 census were included in our original publication plans but were cancelled in budget cutbacks. A private consortium of data users raised the funds needed to prepare the ZIP Code data. The consortium had first access to the data, but the general public eventually had access as well.

Public-Use Microdata Files. Policy analysts, researchers, and data reproducers who want to create their own unique data sets, regression analyses, simulation models, and so forth can do so by using public-use microdata samples (PUMS) that present household sample survey data or census sample data in individual record form. Of course, certain geographic information and respondent characteristics data are suppressed on these files to ensure that the identity of any particular person or housing unit is not disclosed. Yet, they provide enormous detail. For example, a Current Population Survey public-use file may contain age in single years through age 90, occupation coded to more than 500 categories, and earnings in original detail up to an appropriate disclosure limit. The popularity of these files is illustrated by the growth of demand. In 1968, we sold a few dozen public-use files. In 1988, we sold more than a thousand.

Because of technological advances and the amazing power of low-cost microcomputers, it is now becoming feasible and cost effective for an individual to access and use the output of a large survey. A good example is the American Housing Survey. The microdata records from the last biennial survey were recently released on CD-ROM at a cost of \$125. The disk includes about 250 megabytes of data, but it can be used for research purposes on a very low-cost microcomputer set-up. Of course, we are not talking about processing speeds such as one would expect on a minicomputer or mainframe. A CD-ROM is a relatively slow storage device, something like the speed of a floppy disk. Thus, it can take a very long time for a microcomputer to process a large microdata file on CD-ROM, reading and tallying one record at a time; perhaps a number of hours. Such timing is frequently not a problem, and where it is, a user may be able to extract a subset of the data onto a fast hard disk for more acceptable response time.

We are excited about the potential of CD-ROM, and a variety of releases can be anticipated over the near term. We are already looking into CD-ROM releases of the March Current Population Survey for 1988 and 1989, of the National Crime Survey for 1988, and of the American Housing Survey for 1987-88. These likely releases will be accompanied by full documentation (which can be purchased separately for \$5), and we are now conducting CD-ROM user courses at select locations.

Relationship of Machine-Readable Data to Published Data. I have described briefly six ways in which we make data available: printed reports, microfiche, three kinds of machine-readable data files - summary tape files, public-use microdata, and CD-ROM, and special tabulations (which can take any of the foregoing forms).

Although this audience is primarily interested in computerized data, I think it is worthwhile to remind you of the other media for three reasons:

First, some of your data needs can be met more quickly and at a lower cost with information from printed reports or microfiche.

Second, verification that you have accurately retrieved information from summary tapes or microdata files is easy when you have substantial information already tabulated to check against.

Finally, as you process machine-readable data, you frequently need to group data into categories. Review of frequency distributions in existing tabulations may save you several steps in data processing or model estimation.

Economic Programs Data

For the next 18 months, the 1989 Economic and Agricultural Censuses will be our most prolific source of new information. Hundreds of reports, summary tape files, and CD-ROMs will be released over the months ahead. Each of the summary tape files is associated with one or more publications series, so there are always some published numbers for verification purposes.

All computerized releases under the 1989 Economic and Agricultural Censuses programs will be available on CD-ROM; and special ZIP Code files on retail trade, services, and manufacturing will be available only on CD-ROM and computer tape. I should add that these data sets will be the first major release of information that employs the new Metropolitan Statistical Area definitions. The economic data are all summary data and, when properly indexed, response time in normal use can be quite fast, even with large files.

Public-use files have not yet been prepared from economic surveys and censuses because of the unique visibility of establishments, the availability of private sector and regulatory data bases, and the consequent risk of disclosure of confidential data. Demand continues to grow, however, for public-use files on business, particularly those relating to the manufacturing sector. In a conference sponsored by the Census Bureau in 1984, more than 100 economists expressed their desire for such public-use files.

Recently, we developed a longitudinal file of manufacturers called the Longitudinal Research Data Base (LRD). At present, access to these files is limited to staff and selected visiting researchers (NSF-AAA-Census fellow, for example) who must work on-site at the Bureau and be sworn in as special agents of the Census Bureau so they are subject to all the disclosure rules and penalties that apply to the regular staff. By developing models and estimating relationships at the Bureau, however, these researchers are in a position to conduct powerful empirical studies in areas of major policy concern. At present, at least five economic research projects involving researchers from prominent universities are based in part on special LRD estimates prepared at the Bureau. We are exploring a system that permits researchers to submit special requests for computerized statistical analyses and get the results quickly. Also, the staff is working on the development of surrogate public-use files involving data transformations, as a means of

releasing sensitive economic microdata. To be useful, these transformed files must preserve the correct microdata estimates of the economic model; allow the analysis of subsets of the data cross-sectionally and longitudinally; and allow expansion of the file to include new economic variables and the like from outside sources. Two types of transformations are being considered: 1) stochastic transformations that involve adding random noise to the original data while preserving, for example, the mean and variance of the variables and the covariance relationships between variables; and 2) non-stochastic transformations that provide for the release of the transformed release of data in ratio form. Each of these methods has merit, but each limits the types of economic research it will support and in its ability to mask the microdata. We are continuing to work on these knotty problems and expect to find solutions, thereby expanding the range of economic information available to the research community.

Demographic Programs Data

Nowadays, the Census Bureau releases a steady stream of household survey public-use files. Public-use files are edited to remove name, address, and selected geographic identification for smaller places, extreme values for continuous variables, and information that is obtained from or matchable to administrative records systems. There was a time when the user community was fully satisfied with our edited public-use files. That is no longer the case. There are increasing user demands to restore detailed information that was removed from public-use files to protect the identity of survey and census respondents or to add information from administrative records systems to the survey results. Often these demands are clearly in the public interest. That is, the request has a clear-cut federal government program or policy application. However, the Census Bureau's authorizing legislation - Title 13 of U.S. Code - clearly forbids release of data that can be used to identify a particular respondent. Thus, we must be concerned about whether the protections afforded these public-use are sufficient. High speed computers have made public-use files more attractive, but they have also increased public concern about potential abuses to individual privacy resulting from the creation of large integrated data bases. In recent years, events in West Germany, Sweden, and other European countries regarding government data bases have highlighted this concern. Moreover, computer hackers have raised fears that, given enough patience, someone could defeat any scheme designed to protect confidentiality.

With the growing demand for microdata products that cannot be made public under current guidelines and the lack of an acceptable quantitative measure of disclosure risk, the Bureau has undertaken to find solutions that provide our users with the data they want and our respondents with the data protection assurances to which they are entitled. We have established a permanent staff

to study microdata disclosure risk and avoidance. This staff is charged with finding "safe" methods of increasing access to microdata. Some options they are exploring include masking public-use microdata, creating publicly-releasable alternatives to microdata (e.g., tabulated summary statistics by class, correlation matrices of the data, provision of test files so programs can be developed by outside analysts then run by Bureau staff), and special administrative arrangements such as visiting scholar programs.

Our research effort is designed to provide the best possible service to our users - especially federal users - who depend on our data to make policy decisions that affect the quality of life for millions of Americans and are responsible for allocating billions of taxpayer dollars. At the same time, we will avoid the potential risk of identifying survey and census respondents from public-use microdata. Where public-user microdata are not possible given this risk, we will, in time, design alternative products and administrative arrangements, within the requirements of Title 13, that satisfy our users' statistical requirements. We will assure our respondents that the data they provide the Census Bureau for statistical purposes will not be used to make determinations about them as individuals but will be used for the fullest possible appropriate application in the conduct of the public business.

Assisting Data Users

Assisting data users is a major responsibility. We do not release machine-readable files without providing written information about each file. We do conduct training in the use of new technology (e.g., CD-ROM), and we maintain a centralized customer sales/service function.

The Census Bureau has established a substantial reputation for comprehensive documentation, including not only record layout information but glossaries of concept definitions and statements regarding limitations of the data. Documentation is included in the cost of each tape or CD-ROM purchased or can be obtained in advance for a small charge. Acquisition and availability information can be obtained from three census publications:

Major new data files are listed in Census and You, the Bureau's official monthly newsletter for data users. Listings are generally timely, but only selected files are covered.

Monthly Product Announcement (MPA) complements Census and you by giving titles, prices, and ordering information for every publication, microfiche, and data file issued, all with the least possible delay. The MPA does so, however, at the expense of not including any descriptive information beyond the title.

The Bureau of the Census Catalog also presents ordering information, as well as abstracts. Each annual issue covers products issued since 1980.

Our Data User Services Division conducts training programs to help users become acquainted with Bureau products. Of special current interest is the one-day course on CD-ROM. Another component of the user training program is our College Curriculum Support Project, where we develop materials for instructors to use in teaching students about census data. More information on both software and training programs is available from the Data User Services Division.

Our centralized Customer Services Branch in the Data User Services Division handles tape, documentation, and some report sales. Copies of computer tape data files (public-use files and summary tape files) are priced at a standard \$175 per reel of tape, documentation included. The CD-ROM of the American Housing Survey is available at a cost of \$125 as are earlier CD-ROMs, one containing the 1980 Decennial Census ZIP Code file and the other containing the 1982 Census of Agriculture data base by county and the 1982 Census of Retail Establishments by ZIP Code. This summer, we expect to release separate CD-ROMs of geographic data from TIGER (Topologically Integrated Geographic Encoding and Referencing) systems, county business patterns, and the city-county data book. To facilitate distribution, we have developed a system that permits users to establish a deposit account, so they can call in telephone orders and get prompt shipment. Otherwise, it is necessary to send a check or charge to a VISA or Mastercard with your order. The same office serves as a centralized inquiry service on the availability of all Census Bureau data. Its phone number is 301/763-4100. Each of the 12 regional offices of the Census Bureau also handles inquiries, although they do not accept orders.

State Data Center and Clearinghouse Program. Whenever possible, we foster development of downstream data services in the private sector. One of our most successful efforts is the State Data Center program. This cooperative federal-state effort builds on the capacity of state governments to serve the data needs of their constituents. The Census Bureau provides designated state organizations in all of the states with basic data products and training. The state provides staffing and supports a wide variety of data retrieval, promotion, and training activities for users in the state. State Data Centers increasingly are becoming the focal point for local as well as Census statistics. For example, most State Data Centers have vital statistics and other data bases generated within their states. Lists of the State Data Centers and local contact points are available on request from the Data User Services Division.

We also maintain a list of private, public, and academic institutions in our National Clearinghouse for Census Data Services. Each organization registered with the Clearinghouse determines the services it will provide, sets its own fee structure, and is neither franchised nor certified by the Census Bureau. Many organizations on the list provide specialized services, and a few do provide time-shared access to census data on their remote data base.

Evaluation of Data Products

There is one last topic I would like to cover: evaluation of our data products and their uses. Our data users are not usually reticent about letting us know what they want and need. A great many individuals as well as organized groups of users provide excellent feedback; and we have taken steps, many of which I have described here, to improve our products and their accessibility. But we need to know how effective these improvements have been, whether other products warrant improvement, and whether we are spending scarce resources (staff time as well as money) correctly. To answer these questions, we plan to do a Bureau-wide evaluation. This will be a systematic evaluation, perhaps done by an outside contractor, covering all aspects of our data presentation and dissemination programs.. We expect to learn much about how effective the presentation of our myriad data products is, how useful they are to our customers, and what changes ought to be made.

Conclusion

The burgeoning demand for statistical data along with rapid advances in the technology to store and access data present wonderful opportunities for the Bureau of the Census to expand its data production and dissemination programs. We have worked hard to meet increased demands for data by releasing more detailed data products in a variety of formats. By providing data on CD-ROMs, by designing new ways of presenting data such as longitudinal research files, and by developing ways of doing special statistical analyses on request, we are striving to meet user needs. Always mindful of our absolute guarantee to protect the confidentiality of our respondents, we continue to explore new ways to improve access without jeopardizing that trust. □

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.

Data From the Central Bureau of Statistics to the Social Science Community: The Norwegian Experience

by Bjørn Henrichsen¹, *Director
Norwegian Social Science Data Services*

The other presentations in this session are from official producers of statistics. I am representing an organization, the Norwegian Social Science Data Services (NSD), established by the Norwegian Research Council for Science and the Humanities (NAVF) and with strong links to the Norwegian Research Council for Applied Social Research (NORAS).

Due to the composition of the panel, I will, however, first give the background for the Social Science Data Archive movement, and then I will describe briefly our organization and the way we have organized access to data for the research community on Norway. In the presentation I am leaning heavily on the article Stein Rokkan and I wrote about 10 years ago.

NSD was set up by NAVF in 1971 and is to-day financed by the two research councils funding social science research in Norway, NAVF and NORAS. The set-up was a result of two parallel, but still distinctive developments; the one intellectual, the other technological. **Intellectually**, the decisive factor was the increasing prestige of quantitative methods and statistical testing procedures across the entire range of the social sciences; the demand for solid data across large populations spread rapidly from demography and economics, across the entire spectrum of the generalizing disciplines, far into the traditional fields of history and geography.

Technologically, the very notion of a data service was unthinkable before the computer revolution.

In fact the standard definition contrasts the man-readable information stored in the traditional document archive with the machine-readable information stored in the data archive. But the technological conditioning of developments goes much further: the data archive is only one component in the broader set of facilities offered by a data service and all these facilities were developed in one way or another in response to the opportunities opened up by computer technology. A fully-fledged data service

would not only offer files of machine-readable information for various fields of social science research, it would also place at the disposal of the research community a variety of other computer-based tools: facilities for ready conversion of textual data machine-readable form, systems for easy retrieval of information stored in all the files, teaching packages, and workbooks for computer analysis at classroom level, inventories of programs, of on-going research projects, and software systems.

These two sets of developments, the intellectual and the technological ones, interacted at different speeds in different national settings and generated markedly different conditions for institutional innovations.

To gain some insight into the sources of such differences we have to start out from an analysis of the characteristics of what we might call the national "information establishments" on the eve of the computer revolution.

What agencies were there for the collection, storage, classification, and distribution of basic information about social structures and processes?

Institutions for storage and display of information-bearing artifacts came first: the museum, the library, the document archive. Later came the institutionalization of standardized data-gathering operations: tax registers, census, registers of births, marriages and deaths, police records - all elements of a growing system of national book-keeping.

Before the arrival of the computer, these giant data-collection agencies cooperated without great difficulty with the data-storage institutions: the tables and the analyses published by the statistical bureaux were stored in libraries and the original data sheets (census forms, register protocols, and the like) were with some regularity transferred to the established archives.

The computer upset these well-established links: once the central statistical agencies had started to transfer all the new information to magnetic tape or other machine-readable media, there was no stopping the demand for similar technological changes at the storage-display end. The statistical agencies could store all their tapes themselves but this raised serious problems of access for academic social scientist working outside the agencies. The agencies might of course also send the tapes for storage to libraries or archives but this made little sense

as long as these institutions were still firmly wedded to man-readable media: given the predominance of textual material in libraries and archives, it took much longer to introduce even rudimentary computer technology in their operations.

The advent of the computer created a gap in the established network for information flow: at the production end, data were regularly transferred to machine-readable form; at the consumption end, the agencies responsible for storage and distribution to the wider community of social scientists could only handle man-readable data and had no facilities for meeting the new demands.

All this happened just during the period when the methodological revolution in the social sciences had created an increasing demand for mass data and widespread dissatisfaction with procedures of analysis: more and more social scientists had become aware of the potentialities of the mass of information collected by governmental agencies and wanted to subject them to reanalysis.

Theoretically, this problem of data access could have been solved in two ways: the governmental statistical agencies could have set up special divisions for servicing the academic community, or enterprising academics could undertake to reorganize statistics from the central services into manageable data banks for clients in the universities. The first type of solution suited the economists best: they normally tended to get very good service from Central Bureaus. It proved much more difficult to satisfy the academic clientele in the other social sciences: the Central Bureaus were not so easily ready to link up data from censuses, school statistics, elections, and so on to meet the demands of sociologists and political scientists.

This interaction between governmental book-keeping and academic research was heavily influenced by events on another front: the emergence of a methodology offering an alternative to the total enumerations so characteristic of official data collection activities. The demand for facilities for secondary analysis increased markedly in the wake of the methodologies for market research firm and the survey agency proved distinctly more flexible than the established governmental bureaus and were soon able to produce vast quantities of data for the social science community. At first, arrangements for access to such data were irregular and haphazard, but the pressure soon built up for some form of institutionalization.

The result was the establishment of organizationally distinct archives for raw data from polls and surveys, the ICPSR in Ann Arbor, the Roper Center, now in Connecticut and the Zentralarchiv in Cologne are typical examples of this trend.

This development on the narrow "poll-survey" front proved catalytic: it triggered a series of efforts to build up broader-range agencies for the reorganization and transmission of data for research in the social sciences. During the 1960s more and more of the academic data

archives took an interest in the great masses of data available from governmental agencies and started to link up information from many sources into broad-gauged systems for rapid computer retrieval. The Inter-university Consortium in Ann Arbor took an early lead in this direction but a number of groups in Europe had also seen the opportunities. One of these was a group of sociologists and political scientists in Norway: this group had built up its own private files in the early 1960s and was later persuaded to take the step that led to the establishment of the Norwegian Social Science Data Services under the Research Council. This Service covers probably a wider range of different types of data than most of the other archives and deserves some scrutiny just for this reason.

Norwegian social scientists had taken part in the early discussions that led to the establishment of survey archives and had taken steps to acquire the raw data of a number of Gallup polls for reanalysis. But the decisive development took place within the programme of electoral studies launched at the end of the 1950s. Within this programme it proved essential to build up an archive of information for all the communes of Norway: this archive linked up data from election statistics, party membership records, censuses, tax returns and a great variety of other official book-keeping activities. This archive was later expanded into an impressive Commune Data Bank that was to become the core of the Data Service finally established under the Research Council in 1971.

Curiously, the decision to establish a National Data Service was not triggered by increased pressures for access to machine-readable information by mounting dissatisfaction among social scientists with the services they were offered at the University Computing Centers. The transition from mechanical sorting-counting machinery to electronic computers had not improved the conditions for the social science research: on the contrary the many promises of great improvements had caused frustration and anger and the difficulties experienced in writing software specifically for social science users had produced large backlogs of analysis jobs.

By comparison with most of the other data facilities built up over the last two decades, NSD is probably the one giving highest priority to book-keeping and "process produced" data. It is multi-sectoral and sees it as its primary task to link up and to systemize data of different types. We have built close links with the different user groups and have established local offices at all the Norwegian universities and also established close links with data producers like the Central Bureau of Statistics (SSB) and the private polling agencies. From the start the SSB has appointed a member to the NSD Board.

Up until 1976 the SSB responded to requests from individual scholars and from institutions like NSD. Often the same requests came from many users. The procedure was very time consuming both for the SSB and the research community.

In 1976 the SSB was looking for new ways of organizing their transfer of data to the Norwegian research community.

After negotiations between the SSB and NSD, an agreement was signed. The agreement gives NSD a broker function between the SSB and the Social Science Community and it specifies what sorts of data that can be transferred to NSD for further distribution to the individual users.

The agreement secures transfer of both survey data and data from registrations and enumerations. Services to the NSD is in principle free of any charge. Data are transferred to the NSD as soon as data has been processed by the SSB and usually the only restriction on our side is that data shall not be presented before the official publications has been sent out by the SSB.

The purpose of the agreement has been to give the research community as easy access as possible to data from the Central Bureau. Data has been refined by the NSD before they are serviced to the NSD User Community. Both the SSB and the NSD are in their procedures for data transfer following strict rules concerning data protection. The agreement has now been in operation for about 12 years and for the social sciences it has been a success. It has secured easy access to SSB data for the social sciences and stimulated use of such data. In the years since the agreement was signed, the cooperation between the SSB and the NSD has been strengthened and data transfer is to-day secured for most of the areas covered by the Central Bureau. To secure close cooperation the SSB is not only represented on the NSD Board but also on various groups discussing new priorities for the NSD. NSD is also represented in similar groups set up by the SSB to discuss needs and priorities. We are also reporting twice a year to the SSB on the usage of their data, and are sending them publications where their data have been used. As far as I know it is the only agreement of its kind giving free access to such a broad specter of services.

Seen from both the SSB and the NSD I think it is fair to say that the agreement has secured an improvement in the efficiency of the distribution of statistical information in Norway and access has stimulated use of data from the Census Bureau in research at the universities.□

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.

The Role of the Hungarian Central Statistical Office In Providing Social Science Data

by G. Vukovich¹

*Director, Department of Social Statistics
Hungarian Central Statistical Office*

The Central Statistical Office of Hungary - established more than 120 years ago - functions on the basis of two major considerations, or objectives:

-detailed data are collected on a regular basis in physical and monetary terms on the national economy and the social structures and trends.

These data are disseminated in yearbooks, special series and tapes or disks. The Office informs the government and the general public as well as the scientific community both at home and abroad.

The data collected originate partly from the Office's own data collection system and, on the other hand, from the authorities, ministries many of which have their own statistical system for administrative purposes. It is, of course, the Central Statistical Office only which is in the possession of the most important statistical information on all branches of the national economy, which calculates national accounts, national input-output tables, and also has an overview on the most important social trends, the social structure of the nation and the functioning of the social infrastructure.

-Based on the relative richness of information, the Office is preparing scientific, statistical analyses on issues on main interest in the economic and social field.

Most publications do contain such analyses, and the experts working with the Office also publish their papers in different periodicals. The analysis of the data is considered as an important part of the Office's activities. In some cases analyses are made at the special request of the government for decision making purposes.

Social statistics, data on social trends, social phenomena and the social infrastructure are collected in three departments of the Statistical Office:

- the Department of Social Statistics
- the Department for Population Statistics
- the Department of Statistics, on the level of living the latter conducts surveys on household income and expenditure and calculates consumption price indexes.

The Department of Social Statistics is responsible for the information on the social infrastructure (health, education, culture, social security, etc.), for data on different issues of special interest, e.g. crime, alcoholism, accidents, suicides. It is also engaged in data gatherings on the environment focusing also to its human and societal aspects. Another field of its activity is the conducting of sample surveys on special social issues, such as social mobility, time budget and time use, social aspects of housing, lifestyles, the health status of the population, the status of young people, and many others.

The Population Statistics Department is responsible for vital statistics and the census and microcensuses. These data also provide a good deal of information on different social issues.

Social statistics had been rapidly developing during the last fifteen years, and more pronouncedly since 1980.

This was mainly so because many social problems originating from the poor achievement of the economy started to manifest themselves in the midseventies. Until then most existing social anomalies have been latent, hidden behind a growing economy and level of living. However, once the economy began to falter, these problems had started to emerge, became visible and caused ever growing concern in the public and among social scientists.

The government's standpoint had been rather ambiguous at that time: it admitted occasionally the existence of some social problems, like housing shortage but refused to realize the ever growing deepness of social anomalies and behaved rather optimistically concerning the future developments. Several issues, such as the growing poverty, the dysfunction of several social services, the decrease of the living standard, the situation of the old people having minor incomes or pensions, the crisis of the family, were considered as issues of temporary importance, or were not considered at all.

The Statistical Office and some research institutes as well as individual scholars attempted to draw the attention of those responsible for social policies to the unfavourable changes, and these efforts finally resulted in an increasing acceptance of the difficulties, by different Government departments.

Thus the attitude of the official bodies started to change slowly after 1980. The changing atmosphere helped the Social Statistics Department to enlarge the scope of its activities concerning social statistics surveys. A Standardized System of Population Surveys has been set up, consisting of a 2% sample and its subsamples, the latter were used for the surveys themselves. The sampling system is based on the 1980 population census and is being kept up to date by including newly built dwellings. The sampling unit is the household living in the selected dwelling unit.

The setting up of the Standardized System of Population Surveys contributed much to the extension of our scope of activities in special social science fields. It made us possible to conduct a series of surveys which helped scholars and policy makers to discover some burning social problems and to develop a better knowledge of the society and its actual functioning.

It has been clear at the beginning of the eighties that there was a growing need of a deeper analysis of several social phenomena, like the changing way of life, the appearance of the second economy, social mobility, the housing situation, the problems of young people as well as that of the elderly, the causes and consequences of the low fertility, the unfavourable health situation and mortality of the population, the changes in the income distribution, and many others. The surveys provided a great amount of data and analyses and the new findings sooner or later found their way into the general and political public opinion.

From among the surveys which have been carried out during the last years, the most recent one is the time budget and way of life survey taken place in 1986-87. The survey revealed new facts about the growth of the time spent at work in several social groups of the population.

Another important undertaking of the Social Statistical Department has been the morbidity survey carried out in 1986 partly on the same sample as the time budget survey. This provides an opportunity for the analysers for a linkage of the two surveys studying thus the impact of different aspects of life circumstances, social habits, working conditions on the health of the population.

When the plans for the morbidity survey had been worked out we considered the extremely unfavourable mortality conditions of the country and wanted to provide some insight into the mechanism of the causes of deteriorating morbidity. It seems to be more or less proved that overwork, and dietary habits as well as high alcohol consumption are among the causes of the high morbidity and mortality.

The analysis of these interrelations by means of multiple regression is still under way and some of the results will be made available to the public by the end of this year.

The data of the social statistical surveys are regularly stocked in the data bank of TARKI (Social Science Informatic Centre) and made thus available to the

scientific community at home and abroad. Details of the organization of the Social Science Informatic Centre are being discussed at this conference in another paper, we shall therefore not review it here.

Additional forms of dissemination of social science data collected by CSO are the publications and the work done in several scientific committees where sociologists, statisticians and other scholars convene.

In the last ten years several international comparisons were made in the field of time budget-time use, and different types of social indicators. The Hungarian CSO participated and is participating at present at such undertakings with the statistical services of Austria, Czechoslovakia, Finland, Poland, Lettonia and Yugoslavia. □

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.



IASSIST

INTERNATIONAL ASSOCIATION FOR
SOCIAL SCIENCE INFORMATION
SERVICE AND TECHNOLOGY

• • • • •

ASSOCIATION INTERNATIONALE
POUR LES SERVICES ET
TECHNIQUES D'INFORMATION EN
SCIENCES SOCIALES

Membership form

The International Association for Social Science Information Services and Technology (IASSIST) is an international association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from re-

duced fees for attendance at regional and international conferences sponsored by **IASSIST**.

Membership fees are:

Regular Membership: \$20.00 per calendar year.

Student Membership: \$10.00 per calendar year.

Institutional subscriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subscription:

\$35.00 per calendar year (includes one volume of the Quarterly)

I would like to become a member of
IASSIST. Please see my choice below:

- ☐ \$20 Regular Membership
- ☐ \$10 Student Membership
- ☐ \$35 Institutional Membership

My primary Interests are:

- ☐ Archive Services/Administration
- ☐ Data Processing
- ☐ Data Management
- ☐ Research Applications
- ☐ Other (specify) _____

Please make checks payable
to **IASSIST** and Mail to :

Ms Jackie McGee
Treasurer, IASSIST
% Rand Corporation
1700 Main Street
Santa Monica

Name / title

Institutional Affiliation

Mailing Address

City

Country / zip/ postal code / phone

CURRENT RESEARCH

in library & information science

- **CURRENT RESEARCH** is an international quarterly journal offering a unique current awareness service on research and development work in library and information science, archives, documentation and the information aspects of other fields
- **The journal provides** information about a wide range of projects, from expert systems to local user surveys. FLA and doctoral theses, post-doctoral and research-staff work are included
- **Each entry provides** a complete overview of the project, the personnel involved, duration, funding, references, a brief description and a contact name. Full name and subject indexes are included
- **Other features include** a list of student theses and dissertations and a list of funding bodies. Each quarter, an area of research is highlighted in a short article

CURRENT RESEARCH is available on magnetic tape, as well as hard copy, and can be searched online on File 61 (SF=CR) of DIALOG

Subscription: UK £86.00
Overseas (excluding N. America) £103.00
N. America US\$195.00

Write for a free specimen copy to

Sales Department
Library Association Publishing
7 Ridgmount Street
London WC1E 7AE
Tel: 01 636 7543 x 360

LA

LISA

Library & Information Science Abstracts

- **international scope and unrivalled coverage**
LISA provides English-language abstracts of material in over thirty languages. Its serial coverage is unrivalled; 550 titles from 60 countries are regularly included and new titles are frequently added
- **rapidly expanding service which keeps pace with developments**
LISA is now available monthly to provide a faster-breaking service which keeps the user informed of the rapid changes in this field
- **extensive range of non-serial works**
including British Library Research and Development Department reports, conference proceedings and monographs
- **wide subject span**
from special collections and union catalogues to word processing and videotex, publishing and reprography
- **full name and subject indexes provided in each issue**
abstracts are chain-indexed to facilitate highly specific subject searches
- **available in magnetic tape, conventional hard-copy format, online (Dialog file 61) and now on CD-ROM**
Twelve monthly issues and annual index

Subscription: UK £157.00
Overseas (excluding N. America) £188.00
N. America US\$357.00

Write for a free specimen copy to
Sales Department
Library Association Publishing
7 Ridgmount Street
London WC1E 7AE
Tel: 01 636 7543 x 360

LA

P.O. Box # 2138
SANTA MONICA, California
U.S.A. 90406-2138

SERIALS DEPT.
U. OF N. CAROLINA-CHAPEL HILL
CB#3938 DAVIS LIBRARY
CHAPEL HILL NC 27599

LIBRARY

JUL 19 1990

School of Library Science
University of North Carolina